Agents Showing Self-Disclosure. A Preliminary Methodological Approach

Valeria Seidita[a], Angelo Maria Pio Sabella[a] and Antonio Chella[a,b]

[a]Dipartimento di Ingegneria, Universitá degli Studi di Palermo, Italy
[b]ICAR-CNR National Research Council, Palermo, Italy

Abstract

The interaction between humans and robots in Human-Robot Teaming Interaction (HRTI) necessitates robot autonomy, proactivity, and adaptability, as decisions are contingent upon the dynamic context. Trust plays a critical role, and the improvement of human trust and decision-making in robots or agents deployed within such contexts is augmented by robot explainability and self-disclosure. Self-disclosure refers to the ability of robots to effectively communicate pertinent information about themselves, while explainability pertains to the clear communication of robot actions. These concepts are closely interconnected and prove indispensable for effective human-robot interaction. In this paper, we propose a methodological approach for the development of HRTI systems with self-disclosure, leveraging BDI agent technology. The paper delineates how prior efforts in extending the BDI reasoning cycle have contributed to the identification of fundamental design abstractions for HRTI systems and associated agent design activities. This preliminary work underscores the significance of explainability and self-disclosure in human-robot collaboration within HRTI, presenting a pragmatic approach to developing HRTI systems endowed with self-disclosure capabilities through the utilization of BDI agent technology.

Keywords

Self-Disclosure, Explainability, Jason, Agent Oriented Methodology

1. Introduction

Human-Robot Teaming Interaction (HRTI) delves into the effective collaboration between robots and humans to achieve shared goals, encompassing a spectrum of interaction from direct commands to autonomous decision-making. In autonomous roles, robots must not only comprehend the intricacies of the environment and tasks but also allocate responsibilities judiciously. This demands an emphasis on autonomy, proactivity, and adaptivity, given that decisions are inherently context-dependent and dynamically influenced by the ever-changing environment and entities involved [1][2].

In human-only teams, interactions are guided by knowledge of capabilities, interpretation of actions, and trust among team members. Trust plays a pivotal role in task allocation. Knowledge encompasses an understanding of the capabilities of fellow team members, the interpretation of their actions in the context of shared goals, and the level of trust established among teammates.

WOA 2023: 24rd Workshop From Objects to Agents, November 6–8, Rome, Italy
valeria.seidita@unipa.it (V. Seidita); angelomariapio.sabella@community.unipa.it (A. M. P. Sabella); antonio.chella@unipa.it (A. Chella)
0000-0002-0601-6914 (V. Seidita); 0000-0002-4325-759X (A. Chella)

trustworthiness assumes critical importance when determining which actions an entity should autonomously undertake and which can be delegated to others.

Human-Robot Teaming Interaction system hinges on the cooperative efforts of humans and robots working in tandem to execute specific tasks. This involves a mode of communication where human users interact with robots using voice commands, gestures, or user interfaces. In parallel, robots leverage the capabilities bestowed upon them by artificial intelligence and sensors to comprehend the environment and respond appropriately. The overarching goal is to not only bolster the efficiency and safety of operations but also to foster an intuitive interaction between humans and robots.

The Human-Robot Teaming Interaction (HRTI) system aspires to facilitate interaction between humans and robots in the most natural way conceivable. This encompasses a gamut of communication modes, including gestures, natural language, voice commands, and visual communication, all orchestrated to enable human users to communicate intuitively with robots. The intricacies of an HRTI system extend beyond mere communication, incorporating planning and control algorithms that empower robots to perform tasks in concert with humans. These algorithms are meticulously crafted to consider both the capabilities and limitations of the robots and the nuanced preferences and instructions of human users. The integration of HRTI systems with computer systems and software further enables seamless data management and task scheduling, providing a comprehensive framework for automation and control of operations. This multidirectional communication involves human users providing instructions and feedback to robots, while the robots reciprocate by furnishing pertinent information about the status of ongoing activities and operations.

At the core of this intricate web of interactions lies the concept of the team, where each team member contributes not only their individual goals but also the knowledge essential to achieving those objectives. Each team member is entrusted with the responsibility of making reliable decisions regarding which actions to perform autonomously and which to delegate to other members. The augmentation of a robot's capabilities with the inherent ability to explain the rationale behind its actions becomes a pivotal element in elevating its trustworthiness, particularly among human team members, thereby amplifying the quality of interaction within the team.

The notion of "trustworthiness" is deeply entwined with the broader concepts of "explainability" and "self-disclosure". Explainability delves into the robot's capacity to articulate its actions, decisions, and underlying reasoning to humans in a manner that is not only clear but also comprehensible. The ability of a robot to elucidate why it is undertaking a particular action in a language intelligible to humans serves to instill confidence in the decision-making process of the robot. Self-disclosure refers to a robot's proactive communication of relevant information about itself, its capabilities, intentions, and limitations to human team members. Transparent disclosure of limitations or reasons for actions allows humans to make informed decisions about relying on the robot, while the robot uses this information to expand its knowledge base and select actions more efficiently.

Extensively studied in psychology and communication, self-disclosure occurs through various channels, including verbal communication, written communication, nonverbal communication, and online interactions. This practice serves multiple purposes, such as building connections, improving mutual understanding, gaining emotional support, and fostering stronger inter-

personal relationships. In essence, the trio of concepts - trustworthiness, explainability, and self-disclosure - is inherently interconnected in the domain of Human-Robot Interaction (HRI). Navigating this intricate web demands a nuanced methodological approach in the development of systems capable of explicating their actions through self-disclosure.

Human-robot collaboration is increasingly integrated into various aspects of our lives, with potential applications in healthcare, education, entertainment, and the arts. These applications, characterized by the unpredictability and dynamism of the operating environment, necessitate a specific design approach. This paper presents a possible approach for the development of HRTI systems with self-recognition capabilities, drawing from our laboratory's experience in experimenting with and deepening key elements of human-robot interaction. Most experiments have employed agent technology, particularly BDI agent technology [3][4]. The reasoning cycle of a BDI agent, rooted in practical reasoning, aligns well with our objectives. Previous work extended the reasoning cycle to include the ability to justify and explain actions, implemented using JaCaMo and speech acts. This paper identifies basic design abstractions and proposes activities for an agent design methodology by grounding on well known existing agent oriented methodologies..

The paper is organized as follows: Section 2 illustrates previous work, Section 3 illustrates and reviews some existing agent methodologies, Section 4 outlines experiments performed before finalizing the methodological approach, Section 5 outlines our proposal for a preliminary agent design methodology for explainable agents, and finally, Section 6 draws conclusions.

2. Justification and self-disclosure with speech acts

In our preceding research endeavors, our primary focus was directed towards endowing a robot with the capacity to justify its actions, subsequently articulating them aloud - an aspect akin to what is denoted as "inner speech" in the realm of psychology.

The overarching objective of our recent years' research endeavors has revolved around the quest to model and desing agents or robots capable of engaging in interactions founded on trust. The pursuit of this overarching goal encompasses several subgoals: delineating the modeling and implementation processes for agents capable of autonomous decision-making, elucidating the mechanisms for modeling and representing their evolving knowledge during execution, and, finally, exploring how the agent (robot) explicates its own actions.

Our approach has been rooted in an analysis of the cognitive processes preceding the execution of an action, drawing inspiration from corresponding human behavior. We posited that, in the act of choosing an action - instigating the decision-making process - one must first possess a comprehensive model of oneself, one's capabilities aligned with the goal sought in the interaction. Subsequently, we contemplated leveraging the concept of inner speech as a guiding mechanism for the agent in the construction of this model. Inner speech, a pivotal psychological process employed by many individuals in their daily cognitive endeavors for information processing, decision-making, and metacognitive reflections, holds paramount significance. This concept has been subjected to extensive scrutiny across diverse domains, encompassing cognitive psychology, developmental psychology, and cognitive neuroscience [5][6][7]. The concept of inner speech is part of theory of mind and refers to the process of thinking or reflecting using

language in one's own mind without communicating verbally with others. In other words, it is the way people mentally communicate with themselves.

There are various theories of inner speech, but they generally include the following components:

- Inner verbal utterance: people use language in their thoughts, similar to how they would speak aloud, but without making audible sounds. This process of "talking" to oneself in the mind can help with thinking, reasoning, and problem solving.
- Cognitive control: inner speech can play a role in controlling cognitive processes such as self-regulation, planning, and monitoring one's actions. For example, a person might use inner speech to plan a series of steps.
- Reflection and self-awareness: inner speech can be used in reflecting on past, present, or future experiences. It can also be used to evaluate one's actions, feelings, and thoughts.
- Communication with self: inner speech can be used to express and organise complex thoughts. People can use inner speech to process information, solve problems, and make decisions.

In our work we considered and interlaced all these components.

At the beginning of our work, we created a computational model that fits well with the BDI model because it involves a reasoning cycle that begins with the acquisition of the environment and ends in an execution. The reasoning cycle uses various elements as inputs: in addition to inputs from the environment, it also uses inputs from the self, and thus motivation, emotion, and mental states in general. We combined the concept of practical reasoning with a well-known model of trust [8][9] to get the robot to build a model of itself, and included it in the BDI reasoning cycle from a theoretical point of view and in the Jason interpreter cycle in terms of implementation. In the BDI cycle, we extended the deliberation process and knowledge base representation to allow the agent to decompose a plan into a set of actions that are closely related to the knowledge and the agent's capabilities to perform those actions. In this way, agents can maintain a model of themselves and justify the outcome of their actions. We have understood justification to be an essential outcome of self-modelling capabilities.

We chose Jason for the implementation phase because it inherently supports the BDI reasoning cycle. The new computational model could be easily implemented in Jason without having to make significant changes in the agent's programming language. This was a major advantage in terms of maintaining programmers' knowledge of the Jason framework. To implement a Jason agent using our approach, you need to change virtually nothing in the implementation logic. In terms of methodology, in this first phase we experimented with using the TROPOS methodology [10] for requirements analysis and goal breakdown.

In the second part of our work, as mentioned earlier, we experimented with the mechanism of speech acts to realise the ability of self-disclosure through the so-called inner speech. Speech acts realise how agents act during communications. The basic principle of speech act theory lies in the meaning of speech. The principle can be summarised by assuming speech as an act [11][12].

This approach allows us to consider the actions encoded in the plans. In Figure 1 the extended part of the algorithm for including both justification and inner speech. For each action, starting

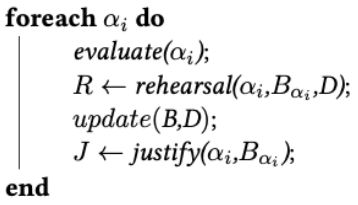

Figure 1: Extension in the BDI reasoning cycle.

from the beliefs about that action and its goals, a function is initiated and then implemented, it produces what we have called rehearsal. Rehearsal is a concept closely related to the concept of the inner sphere. It allows the realisation of feedback that the agent externalises and is also reread (or heard) by the same agent so that it can modify its knowledge base (the *update(B,D)* function) and so to justify its actions.

In this first phase, we have focused on rehearsal which produces justification and an increase in knowledge, but does not yet have an impact on the thought process. At the moment, we have an indirect effect on the thinking process through the direct change in the agent's knowledge of the environment and others as a result of the speech act.

The examination of the psychological dimensions and confidence-building implications associated with an agent's capacity to articulate self-explanations has constituted a focal point in our laboratory, approached through a non-agent paradigm. As delineated in [13][14], our studies have demonstrated a significant augmentation in human trust towards a robot endowed with self-explanatory capabilities. Notably, the generation of internal discourse was a little static to meet requests by the psychology team to have feedback consistent for certain interactions..

However, the latter part of this work allowed an exploration into agent technology for the autonomous generation of inner speech and justifications. Building upon the initial findings, our ongoing efforts involve refining the methodological underpinnings of this approach, concurrently serving as an additional layer of validation for our research endeavors. In essence, our current focus involves the simultaneous refinement of both conceptual and methodological dimensions.

3. BDI design methodology and their relevant features

Belief-Desire-Intention (BDI) agent-based methodologies are a fundamental approach in Artificial Intelligence for designing multi-agent systems. These methodologies are based on an agent model that represents the beliefs, desires, and intentions of agents and enables them to make rational decisions in a complex environment. In our work work we analysed several BDI agent methodologies and this section, we briefly review a some of them making considerations to outline a new design methodology.

The GAIA methodology [15] represents a significant approach in the field of multiagent systems that builds on the basic principles of the Belief-Desire-Intention (BDI) model. In GAIA, the concepts of Belief, Desire, and Intention are used to design the reasoning cycle of agents.

In GAIA, “Beliefs” represent agents’ shared knowledge about the environment, “Desires” are the goals to be achieved, and “Intentions” are the concrete actions to fulfil these goals. The reasoning cycle of agents in GAIA includes the phases of perception, belief updating, evaluation of desires, generation of intentions, selection of intentions, and execution. These BDI concepts guide agent behaviour in GAIA and provide a solid foundation for developing intelligent multi-agent systems in various applications. In the context of GAIA methodology, the concept of “organization” refers to the structure or framework that defines how agents are organized, interact, and cooperate within a multiagent system. Organization is a fundamental aspect of multiagent system development, and its design influences the overall dynamics and effectiveness of the system. Organization in GAIA defines the relationships and interactions among agents in the environment. This can include hierarchies, networks, roles, and communication rules that govern how agents work together. In designing the organization, specific roles are defined and agents can play. them These roles can be assigned based on the capabilities of the agents or the needs of the system. In our proposed methodology, this aspect is strongly considered, except for the concept of team that is used instead of organization.

SODA is a methodology [16][17] that focuses on the design of agent societies in open, distributed environments. This framework uses an agent society-based perspective to model multi-agent systems and foster collaboration among autonomous agents in complex contexts. SODA extends the basic concepts of the BDI model (Belief, Desire, Intention) with additional elements, including “goals” and a focus on the environment in which agents operate. “Goals” represent the objectives or targets that agents seek to achieve. Goals are an essential component of agent decision making in SODA, in addition to Belief, Desire, and Intention. Agents constantly evaluate goals based on their beliefs and desires to decide which goals to pursue. Beliefs represent agents’ knowledge about the environment in which they operate. These beliefs influence the evaluation of goals and the formulation of agents’ intentions. “Desires” are the goals or desires of the agents. Agents continuously evaluate their desires based on current beliefs and the needs of the environment. “Intentions” in SODA represent the concrete actions that agents formulate to fulfill their desires and achieve goals. They are the final step in the agents’ decision-making process before acting. The concept of “environment” in SODA is essential. It represents the context in which agents act and can be complex and distributed. The design of the environment is crucial in SODA because it influences the agents’ beliefs and opportunities to pursue goals. The environment can include resources, obstacles, other agents, and shared data.

In the same way we analysed INGENIAS [18], JACK [19] and then TROPOS [10] for the part regarding the goal-oriented analysis and were able to outline the methodology presented in the following section.

4. From the case study towards design abstractions

The utilized scenario in our preceding research constitutes a straightforward collaborative setting, focusing on the shared objective of a team comprising humans and robots - arranging a table. The environment is equipped with all requisite objects, detailed mission instructions delineating the actions to be executed, and the spatial arrangement of objects. The adherence to specific rules governs the assembly of dishes on the table, thereby determining the actions of

team members. Each agent is tasked with the selection of actions and the designated object for their execution. Ideally, this decision-making process operates through an adoption-delegation mechanism, where each member leverages their knowledge and observes others' actions or anticipates their intentions.

The incorporation of rules and etiquette significantly influences the agents' behavior and necessitates meticulous consideration during the design phase. As previously alluded to, the implementation of this mechanism is realized through the characteristics of Jason agents [20][21], with the environment modeled using CArtAgO [22][23]. Furthermore, an extension of the reasoning cycle, as detailed in our prior work and expounded upon in the preceding section, complements these foundational components. This scenario has been used in the projects we have done in the robotics lab in recent years. It was mainly used to study the psychological aspects of trust in the robot when a robot can express aloud what it is doing. The scenario involves a robot and a human. In the design phase, the robot (and the agents) are given a set of plans with which to set the table. In the execution phase, the robot determines the action it thinks is best and uses speech acts and justifications to explain why it chose certain actions and/or their outcome.

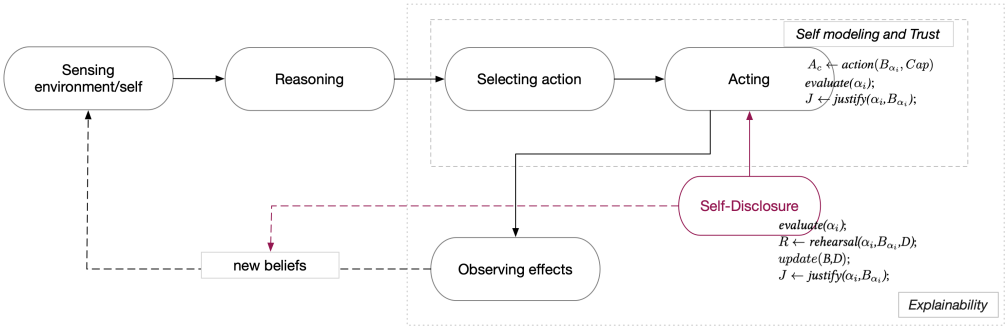

Figure 2: The process for self-discloure with inner speech (reported by [24]).

Figure 2 illustrates the process that connects the modules within the architecture discussed in [24]. Agents perform reasoning processes analysing inputs from their environment, which include both external and internal stimuli. They select actions from a predefined set provided by the designer. Once an action is selected, it observes the consequences of its actions and the resulting changes in the external world. it updates its beliefs through perception and activates the justification function. Justification is primarily concerned with providing an explanation for the action that focuses exclusively on the results achieved.

The next step is to introduce new beliefs or modify existing ones to account for the effects of inner speech, which essentially reflects what the agent thinks. For each selected action, the agent evaluates whether it can be performed by assessing the pre- and post-conditions of the action in light of its knowledge base. Following this evaluation, the agent revises its beliefs and desires through the rehearsal function.

Observing events via inner dialog influences the agent's mental state, resulting in new desires and an updated set of beliefs. A BDI agent takes a similar approach by using communication with other agents to influence changes in their beliefs. In this particular scenario, the agent

sends a message to itself as a result of the rehearsal process.

The design methodology we propose is the result of an iterative and gradual work that included first a practical and then a theoretical approach. Starting from experiments with the technology and the elements that can be implemented in Jason and CArtAgO, we then identified the main abstractions on which the design process is then based. Jason and CArtAgO are based on a very specific metamodel (shown in Figure 3) in its variant, which shows which elements intervene and are fundamental to the design's self-disclosure in the form of the inner speech.

Moreover, the application context is very specific, as it is not so much a human-robot interaction, but a human-robot interaction in teams. In a team, the goal is shared. Even if it were broken down, it would have no purpose other than to identify actions that then need to be reassembled in the form of plans to achieve the goal. Identifying the goal is the first thing we tackled in implementing this scenario, albeit a simple one, but it applies to the more complex goals as well. After that, the goals are divided into sub-goals and tasks are assigned to each goal, and then the role that an agent should play is determined. So first the role is identified, the role is assigned a task, and then the agent is identified to perform that task.

The concept of "role" in agent design methodologies is a key element that helps define an agent's behaviour and responsibilities in a multi-agent system. This concept is used to organise and structure the architecture and behaviour of agents within a system and to provide a clear division of activities and functions.

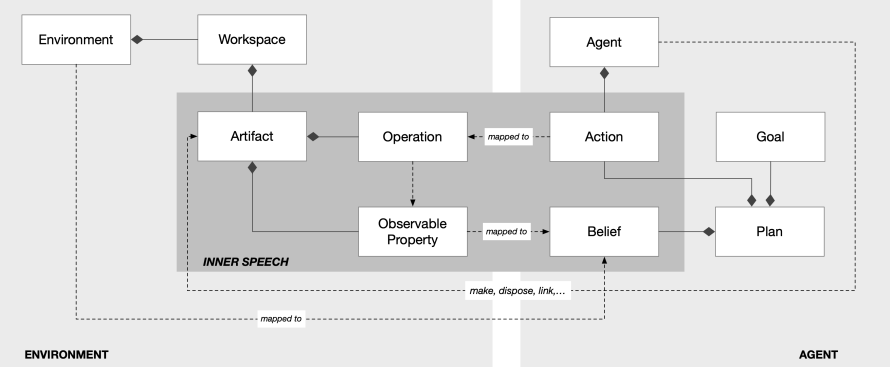

Figure 3: Jason and CArtAgO metamodel and the relations with elements realising the inner speech (reported in [24]).

In the approach we propose, the role concept is primary in establishing connections with other design elements for reasons we will discuss below. Each agent within a multiagent system is assigned one or more specific roles. A role represents a set of responsibilities and activities that an agent must perform in the context of the application. For example, in a traffic management system, an agent might have the role of "traffic controller" and would thus be responsible for controlling the flow of traffic at an intersection. In terms of collaboration, agents with different roles can work together to achieve common goals. For example, the agent in the example above could collaborate with the agent with the role "monitoring road conditions" to make informed traffic management decisions. In a team, agents can be assigned to different roles or take on new roles in response to changes in the environment or application requirements, making them

flexible and adaptable.

Roles facilitate coordination between agents within the system. Agents with complementary roles can coordinate their actions to achieve the overall goals of the system.

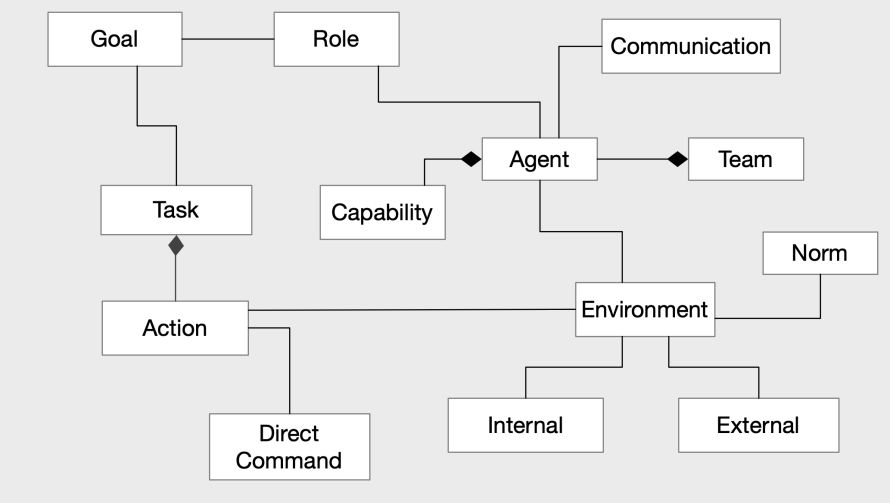

Figure 4: The main metamodel for designing self-disclosure agents.

In Figure 4 the analysis of the basic design elements that emerge from the case study investigation is summarised in the form of a metamodel. This metamodel represents an higher level of abstraction with respect to the model in 3.

5. A first proposal of design methodology for agents explaining themselves

The purpose of this paper is to outline an initial hypothesis of an agent-based design methodology for the development and implementation of an explainable human-robot teaming interaction system by the implementation of self-disclosure capabilities. As presented in the previous sections, the application context has very unique characteristics compared to the complex problems normally handled in the multi-agent environment. For this reason, an ad hoc design methodology is required. The main difference lies in the concept of the team and the goal. In the main agent methodologies studied, the goal is something that can be achieved as a composition of subgoals that can be assigned to different agents to bring about the desired solution, and the agents can be organised into an organisation that regulates and structures the behaviour of the agents in some way. However, in the context we are considering, the central element is the team, which has only one shared goal. The subdivision of the goal into subgoals serves either to simplify the problem in order to assign individual parts to different teams, or to identify atomic goals, which are then assigned actions that are part of a plan. The peer rather than hierarchical view and this implies an adaptation of theories that are below the known methodologies.

We considered adapting, extending, and merging existing methods because the BDI agent technology, the Jason programming language, and the CArtAgO framework proved efficient in managing resources, knowledge representation, and plans during initial experiments.

Jason and CArtAgO are two languages and development platforms widely used in the development of multi-agent systems based on cognitive agents. Below are some of the key design abstractions for designing a system using Jason and CArtAgO.

Jason is based on the BDI model, which represents agent behaviour through plans and rules, in addition to beliefs, desires, and intentions. Jason agents follow action plans defined by user-designed rules. The rules define how the agent should react to available information (beliefs) and changes in environmental conditions. Communication between agents is also an important aspect. Jason agents can send messages to other agents to exchange information, coordinate actions, and negotiate goals. Finally, Jason supports distributed reasoning, allowing agents to process information autonomously and make decisions based on their knowledge and goals.

CArtAgO focuses on the environment and perception. CArtAgO is a development platform for physical agents and one of the basic abstractions is the environment in which the agent operates. It is necessary to define how the agent perceives its environment, including sensors and perception data. In CArtAgO, agents perform actions to achieve specific tasks or goals. It is important to define what tasks the agent must perform and how the actions are performed to accomplish those tasks. As with most agent methodologies, and based on the experience in the implementation phase of the case study in section 4, the abstractions of tasks and actions are critical to agent control. CArtAgO allows agents to interact directly with the physical world via actuators. This abstraction is important for developing physical agents that perform actions in the real world. Finally, the concept of artifact is important in CArtAgO. It is a key abstraction for managing resources and agent interaction. These resources can include physical objects, data, services, and more. Artifacts allow agents to perceive, manipulate, and use resources in the environment. Agents can access artifacts to obtain information or influence the state of the resources themselves. This can include obtaining data, modifying physical objects, or using services provided by artifacts. Artifacts also serve as a point of communication between agents. Agents can use artifacts to share information and coordinate their actions. This supports collaboration and knowledge sharing between agents.

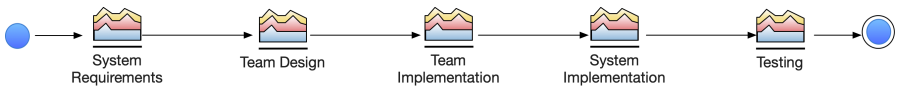

Figure 5: The desing methodology phases

Artifacts along with beliefs allow for easy and efficient implementation of metamodel elements that relate to the environment, both externally and internally.

Based on what has been shown so far, Figure 5 shows the phases of our proposed design methodology. It consists of five phases, each of which is divided into activities, as shown in the following figures. The phases are:

- System Requirements - Understanding and modelling the system requirements and their

focused description along with system knowledge analysis;

- Team Design - In this phase, all aspects of the team of agents that will later be deployed in the robotic system are identified and modelled. The roles and tasks assigned to them, as well as the knowledge required to perform each task, are identified and described;
- Team implementation - the transition from the previous phase to the technical aspects: Artifacts, Plans, Rules.
- System implementation - is the phase where all the details about the use of the agents in the robotic system are analysed.
- Testing - evaluation and verification that the interactions are performed as intended.

All the process is performed by the transformation of models leading from the abstraction level of Figure 4 to that of Figure 3 and then to the actual code.

Figure 6 and 7 show a detail of the activities in the first two phases that are the focus of this paper. Each activity is accompanied by a work product and in each of them all elements of the metamodel from Figure 4 are instantiated and described. So far, for each activity, methods from other methodologies and on the base of the experience of the RoboticsLab software engineering team at the University of Palermo have been used and adapted. A more detailed description following the "IEEE-FIPA Standard on the Design Process Documentation Template" [25] will be the subject of further work.

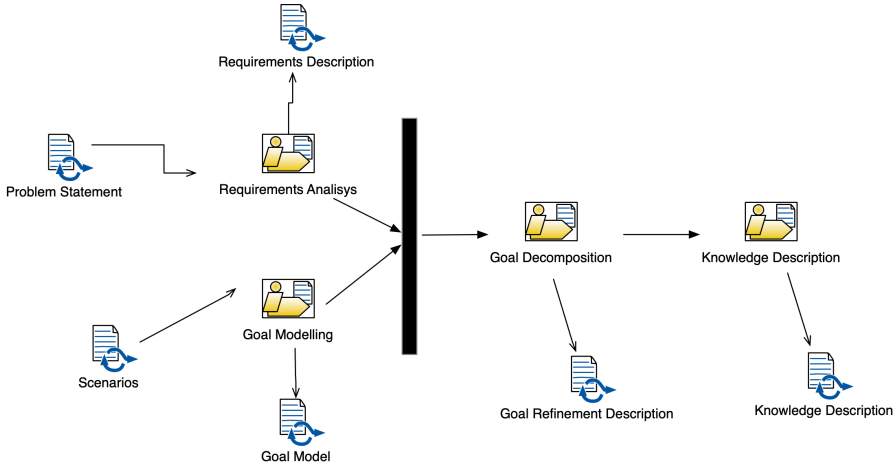

Figure 6: The System Requirements Activities.

6. Conclusion

Adequate explainability helps improve perceptions of trustworthiness, as humans are able to understand robots' actions and decisions. At the same time, self-disclosure helps build trust by providing clear information about the robot's capabilities and limitations. When all three factors are managed well, the result is a collaborative and safe working environment in which humans and robots can work together effectively to achieve common goals.

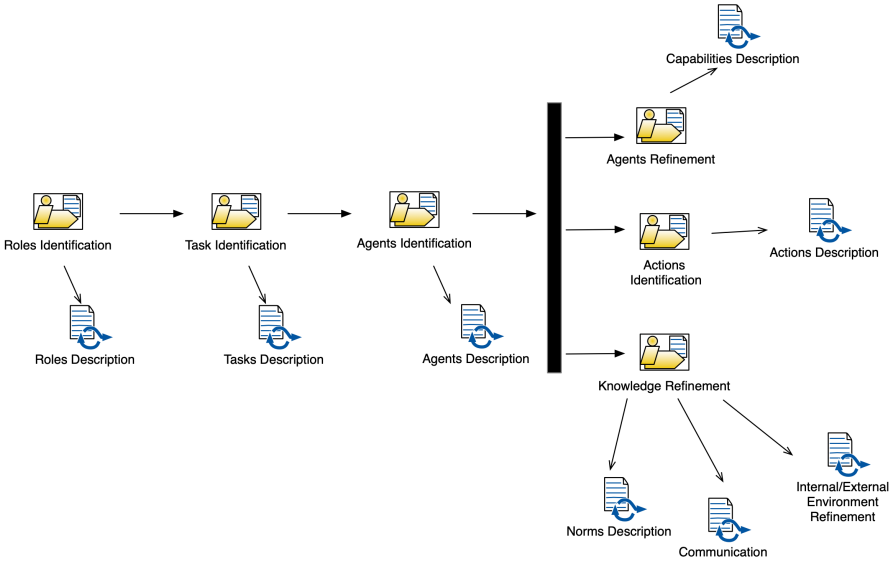

Figure 7: The Team Description Activities.

In the context of human-robot interaction, it is important to develop deployable systems that can interact with team members in ways similar to humans. Agent technology is a very powerful tool to implement these types of systems, but a design problem arises. Existing methods cannot be used to their full potential because they use abstractions that are not a perfect fit for HRTI.

We have taken advantage of previous experiences in building trustworthy HRTI systems that explain our own behaviour, and have abstracted the elements of a design methodology by hypothesising some of the key phases and activities.

The role concept in agent design methodologies has been an important mechanism for structuring and organising agent behaviour within a multi-agent system. It provides a clear definition of responsibilities, encourages specialisation, promotes collaboration, and makes the system more flexible and efficient. Role design is a critical step in building robust and scalable agent systems. In addition, we considered that both Jason and CArtAgO provide specific abstractions for designing agents and multi-agent systems. The choice between the two depends on the type of system to be developed and the specific requirements of the project. Both provide powerful tools for the design and implementation of complex multiagent systems.

The result obtained and illustrated in this paper is the result of using a tabletop scenario. The scenario is very simple, but certainly valid for a first hypothesis of the methodology. In the future, we will repeat and refine our work by using and developing more complex scenarios and detailing the individual methods within each activity.

References

[1] L. Mingyue Ma, T. Fong, M. J. Micire, Y. K. Kim, K. Feigh, Human-robot teaming: Concepts and components for design, in: Field and Service Robotics: Results of the 11th International

Conference, Springer, 2018, pp. 649–663.
[2] A. Chella, F. Lanza, V. Seidita, A cognitive architecture for human-robot teaming interaction, in: Proceedings of the 6th International Workshop on Artificial Intelligence and Cognition, Palermo, 2018.
[3] A. S. Rao, M. P. Georgeff, et al., Bdi agents: from theory to practice., in: Icmas, volume 95, 1995, pp. 312–319.
[4] L. De Silva, F. R. Meneguzzi, B. Logan, Bdi agent architectures: A survey, in: Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI), 2020, Japão., 2020.
[5] L. S. Vygotsky, Thought and language, MIT press, 2012.
[6] A. Morin, When inner speech and imagined interactions meet, Imagination, Cognition and Personality 39 (2020) 374–385.
[7] A. Morin, Possible links between self-awareness and inner speech theoretical background, underlying mechanisms, and empirical evidence, Journal of Consciousness Studies 12 (2005) 115–134.
[8] R. Falcone, C. Castelfranchi, Social trust: A cognitive approach, Trust and deception in virtual societies (2001) 55–90.
[9] C. Castelfranchi, R. Falcone, Towards a theory of delegation for agent-based systems, Robotics and Autonomous systems 24 (1998) 141–157.
[10] P. Bresciani, A. Perini, P. Giorgini, F. Giunchiglia, J. Mylopoulos, Tropos: An agent-oriented software development methodology, Autonomous Agents and Multi-Agent Systems 8 (2004) 203–236.
[11] J. Searle, J. R. Searle, Speech acts: An essay in the philosophy of language, volume 626, Cambridge university press, 1969.
[12] J. Austin, How to do things with words, Oxford university press, 1975.
[13] A. Pipitone, A. Geraci, A. D'Amico, V. Seidita, A. Chella, Robotś inner speech effects on trust and anthropomorphic cues in human-robot cooperation, arXiv preprint arXiv:2109.09388 (2021).
[14] A. Geraci, A. D'Amico, A. Pipitone, V. Seidita, A. Chella, Automation inner speech as an anthropomorphic feature affecting human trust: Current issues and future directions, Frontiers in Robotics and AI 8 (2021) 620026.
[15] M. Wooldridge, N. R. Jennings, D. Kinny, The gaia methodology for agent-oriented analysis and design, Autonomous Agents and multi-agent systems 3 (2000) 285–312.
[16] A. Omicini, Soda: Societies and infrastructures in the analysis and design of agent-based systems, in: International workshop on agent-oriented software engineering, Springer, 2000, pp. 185–193.
[17] A. Molesini, A. Omicini, The soda methodology: Meta-model and process documentation, Handbook on Agent-Oriented Design Processes (2014) 407–461.
[18] J. Pavón, J. J. Gómez-Sanz, R. Fuentes, The ingenias methodology and tools, in: Agent-oriented methodologies, IGI Global, 2005, pp. 236–276.
[19] M. Winikoff, Jack? intelligent agents: an industrial strength platform, Multi-agent programming: Languages, platforms and applications (2005) 175–193.
[20] A. S. Rao, Agentspeak (l): Bdi agents speak out in a logical computable language, in: European workshop on modelling autonomous agents in a multi-agent world, Springer,

1996, pp. 42–55.
[21] R. H. Bordini, J. F. Hübner, Bdi agent programming in agentspeak using jason, in: International workshop on computational logic in multi-agent systems, Springer, 2005, pp. 143–164.
[22] D. Weyns, A. Omicini, J. Odell, Environment as a first class abstraction in multiagent systems, Autonomous agents and multi-agent systems 14 (2007) 5–30.
[23] A. Omicini, A. Ricci, M. Viroli, Artifacts in the a&a meta-model for multi-agent systems, Autonomous agents and multi-agent systems 17 (2008) 432–456.
[24] V. Seidita, A. M. P. Sabella, F. Lanza, A. Chella, Agent talks about itself: an implementation using jason, cartago and speech acts, Intelligenza Artificiale 17 (2023) 7–18.
[25] M. Cossentino, V. Hilaire, A. Molesini, V. Seidita, The ieee-fipa standard on the design process documentation template, Handbook on Agent-Oriented Design Processes (2014) 7–17.